

National Center for Biotechnology Information Resources for Plant Genomics

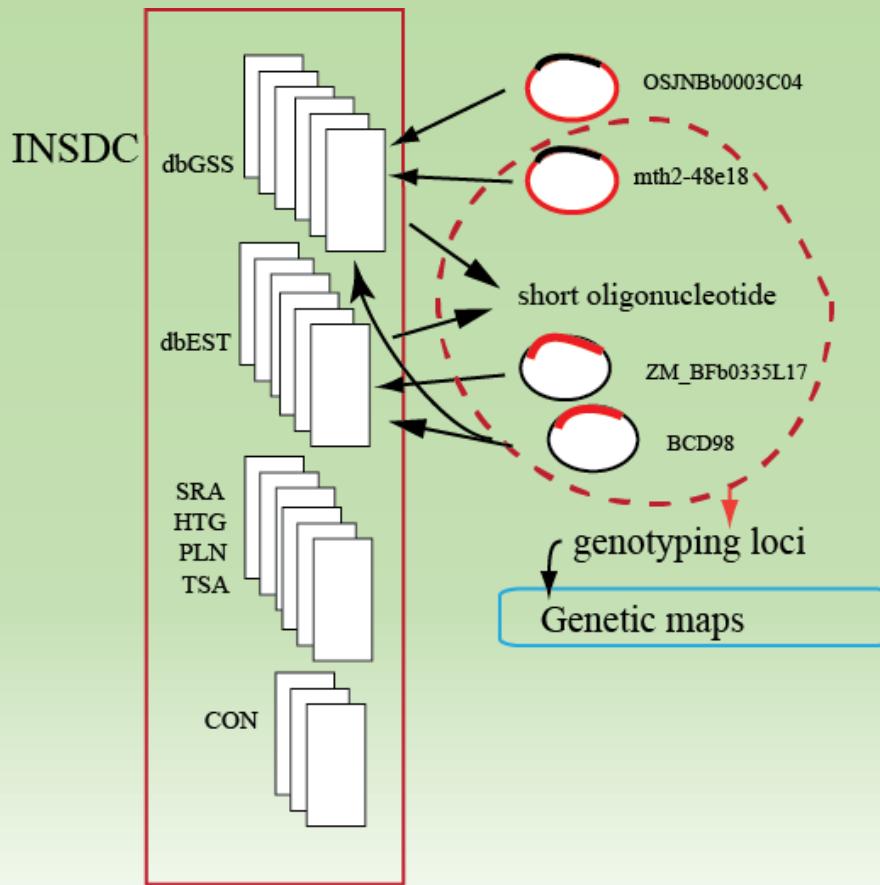
Brian Smith-White

Tuesday 13 January 2015

Acknowledgements

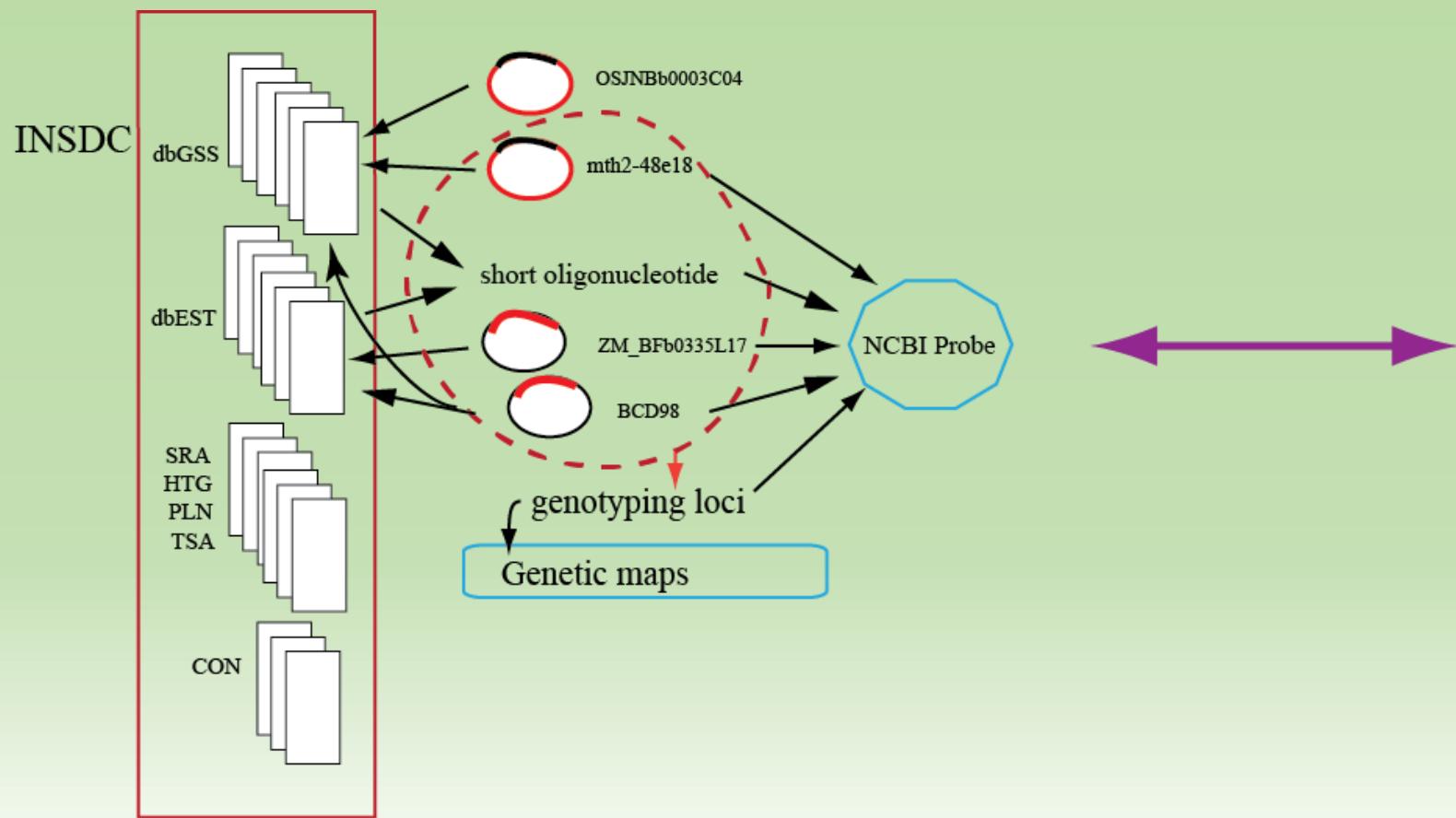
- Svetlana Zazvovskaia – NCBI Probe group
- Anjana Raina – RefSeq group
- Terence D. Murphy – NCBI Gene group
- Ilene Mizrachi – NCBI Probe group
- Valerie Schneider – CloneDB group
- Kim D. Pruitt – RefSeq group

Nucleic acid reagents



NCBI Probe -

<http://www.ncbi.nlm.nih.gov/probe>



Granularity of NCBI Probe data – same name, different techniques

NCBI Resources How To

Probe Save search Limits Advanced

Display Settings: Send to:

Results: 3

[STS probe TG174](#)
1. Accession: Pr012790551 ID: 12790551
 Name: TG174
 Type: STS

[Solanum lycopersicum CAPS probe TG174](#)
2. Accession: Pr006097453 ID: 6097453
 Name: TG174
 Type: CAPS
 Application: genotyping
 Target organism: [Solanum lycopersicum](#)

[Solanum lycopersicum RFLP probe TG174](#)
3. Accession: Pr02246415 ID: 2246415
 Name: TG174
 Type: RFLP
 Application: genotyping
 Source organism: [Solanum lycopersicum](#)
 Target organisms: [Capsicum annuum](#); [Solanum lycopersicum](#); [Solanum peruvianum](#); [Solanum pimpinellifolium](#); [Solanum pennelli](#); [Solanum habrochaites](#); [Capsicum chinense](#); [Capsicum annuum x Capsicum chinense](#)

Granularity of NCBI Probe data – same name, different organism

NCBI Resources How To

Probe Save search Limits Advanced

[Display Settings:](#) Summary [Send to:](#)

Results: 3

[STS probe TG174](#)
1. Accession: Pr012790551 ID: 12790551
Name: TG174
Type: STS

[Solanum lycopersicum CAPS probe TG174](#)
2. Accession: Pr006097453 ID: 6097453
Name: TG174
Type: CAPS
Application: genotyping
Target organism: [Solanum lycopersicum](#)

[Solanum lycopersicum RFLP probe TG174](#)
3. Accession: Pr02246415 ID: 2246415
Name: TG174
Type: RFLP
Application: genotyping
Source organism: [Solanum lycopersicum](#)
Target organisms: [Capsicum annuum](#); [Solanum lycopersicum](#); [Solanum habrochaites](#); [Capsicum chinense](#); [Capsicum annuum x Capsicum chinense](#)

NCBI Resources How To

Probe Save search Limits Advanced

[Display Settings:](#) Summary [Send to:](#)

Results: 2

[Hordeum vulgare RFLP probe pA086 for cold-regulated mRNA \(M60733\)](#)
1. Accession: Pr004066840 ID: 4066840
Name: pA086
Type: RFLP
Application: genotyping
Source organism: [Hordeum vulgare](#)
Source sequence: [M60733](#)
Target organism: [Hordeum vulgare](#)

[Glycine max RFLP probe pA086 for T134012 Soybean RFLP probe Glycine max genomic clone A086 \(AQ841762\) and 3 more sequences](#)
2. [sequences](#)
Accession: Pr003647552 ID: 3647552
Name: pA086
Type: RFLP
Application: genotyping
Source organism: [Glycine max](#)
Source sequences: [4 sequences](#)
Target organisms: [Glycine max](#); [Glycine soja](#)

Display of synonym data

Probe name display

NCBI Resources > How To >

Probe Save search | Limits | Advanced

Display Settings: Full Report
P001545

Hordeum vulgare RFLP probe pHTB1 for vacuolar ATPase B subunit mRNA (L11862)

Synopsis

| FieldName | Values |
|------------------|--|
| Name | pHTB1 |
| Aliases | Vmp-B1, Vmp57A, VMp57A |
| Type | RFLP |
| Application | genotyping |
| Source organism | Hordeum vulgare |
| Source sequence | L11862 (Hordeum vulgare) |
| Target organisms | Triticum monococcum vulgare_monsacrum; Hordeum vulgare; Triticum monosaccharum; Triticum monococcum subsp. aestivoides |

Sequences

Synonym name display

NCBI - Probes ID: How To >

Probe Save | Advanced

Display Settings: Full Report
P003246008

RFLP probe Vmp57A

Synopsis

| FieldName | Values |
|-----------------|------------|
| Name | Vmp57A |
| Type | RFLP |
| Application | genotyping |
| Source organism | |
| Source sequence | |

Sequences

No Sequence Data.

References

1. A Klineholz via Gramene 2009
2. A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Kishiiwa A, Kilan A, Sagha Marsh MA, Blyashen RM, Hayes P, Chen FG, Lapitan N, Fennock A, Blaak TX, Karaszi V, Aranee E, Dahlman L, Kadria D, Bellinger J, Krupp SJ, Liu H, Saneeli M, Heun M, Franckowiak JD, Hoffman D, Staden R, Thielman BJ. Theoretical Applied Genetics. 1982;66:761-712.

Additional Data

| FieldName | Values |
|-----------------|--|
| Probe Cross ref | dbProbes_1646 |
| Submission Info | Journal list of probes in genetic maps |

Display of synonym data – multiple probes with identical synonym

The screenshot shows the NCBI Probe search interface. A blue arrow points to the left sidebar, which contains links for Probe, Display Settings, Ph001545, Hordeum vulgare, Synopsis, Field Name, Name, Aliases, Type, Application, Source organism, Source taxon, Target organism, and Sequences. The main search bar has 'Probe' selected and 'AFLP11' entered. Below the search bar are buttons for 'Save search', 'Limits', and 'Advanced'. The search results section displays four entries, each with a checkbox and a detailed description. To the right of the results are sections for 'Send to:' (with a dropdown menu showing 'All (4)' highlighted in yellow), 'Filter your results:' (with options for Variation, Silencing, and Expression), 'Results by taxon' (listing Top Organisms: Avena strigosa x Avena wiestii (1) and Theobroma cacao (1)), 'Find related data', 'Search details', and 'Recent activity'. The bottom of the page includes 'Display Settings' and 'Send to:' buttons.

NCBI Resources How To Sign in to NCBI

Probe AFLP11 Help

Display Settings: Summary

Send to: Filter your results:

All (4)

Variation (0)

Silencing (0)

Expression (0)

Manage Filters

Results: 4

[AFLP probe set AFLP11](#)
1. Accession: Pr031976813 ID: 31976813
Name: AFLP11
Type: AFLP

[Theobroma cacao AFLP probe set E40M6062](#)
2. Accession: Pr031905623 ID: 31905623
Name: E40M6062
Alias: AFLP11
Type: AFLP
Target organism: [Theobroma cacao](#)

[AFLP probe set AFLP11](#)
3. Accession: Pr031899717 ID: 31899717
Name: AFLP11
Type: AFLP

[Avena strigosa x Avena wiestii AFLP probe set E34M56](#)
4. Accession: Pr031870007 ID: 31870007
Name: E34M56
Alias: AFLP11
Type: AFLP
Target organism: [Avena strigosa x Avena wiestii](#)

Display Settings: Summary

Send to:

Results by taxon

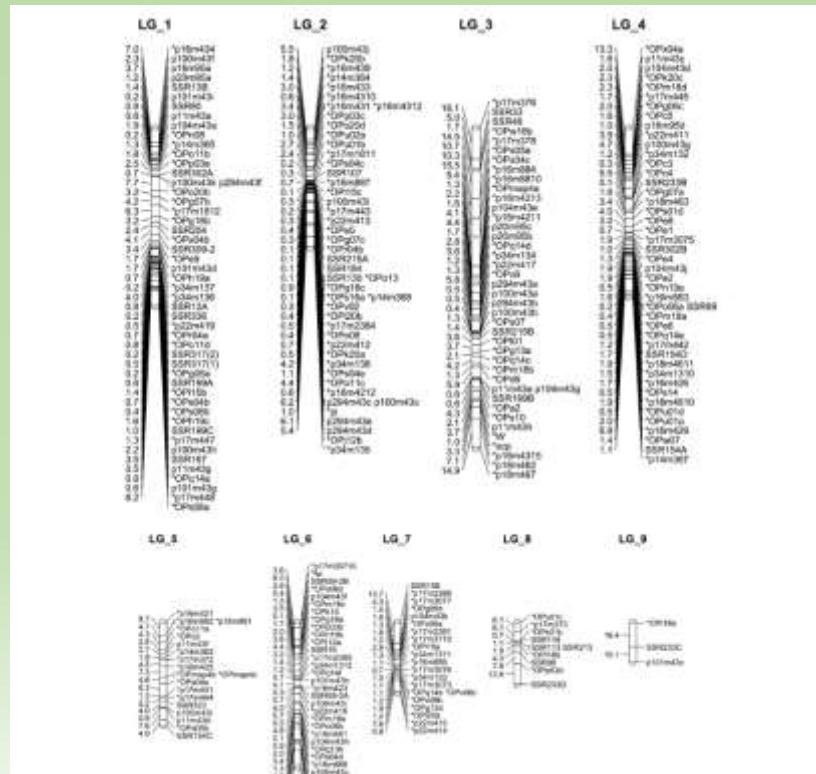
Top Organisms [Tree]
Avena strigosa x Avena wiestii (1)
Theobroma cacao (1)

Find related data

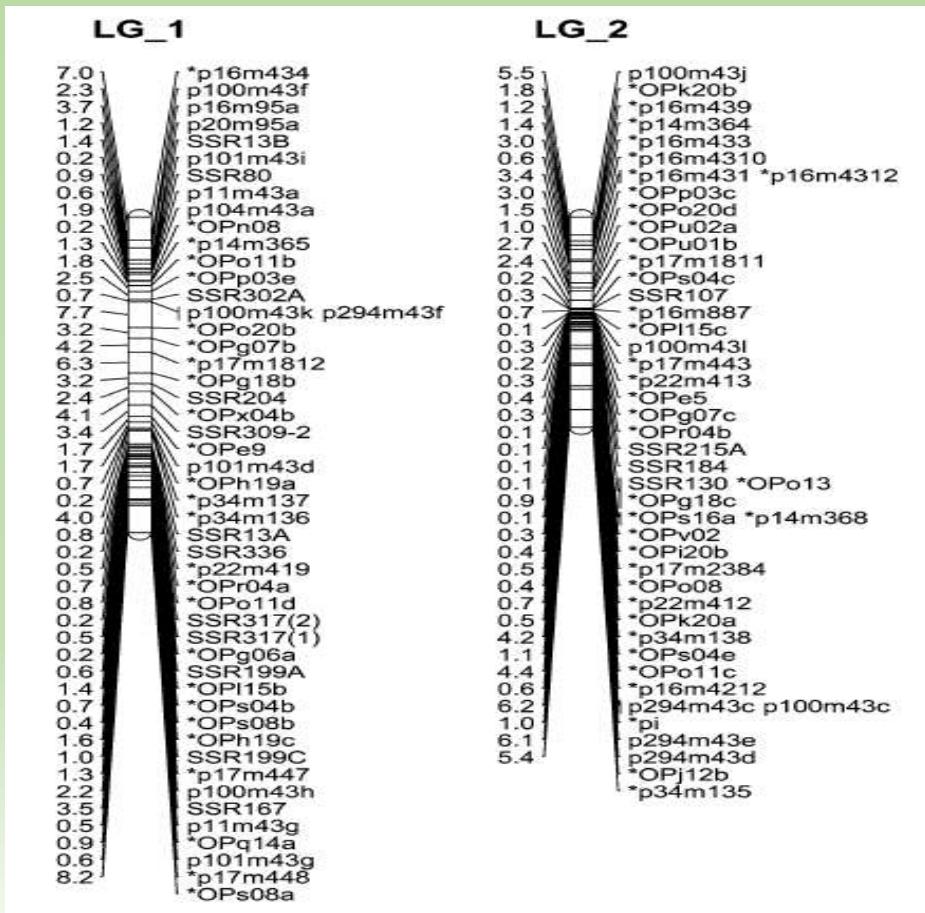
Search details

Recent activity

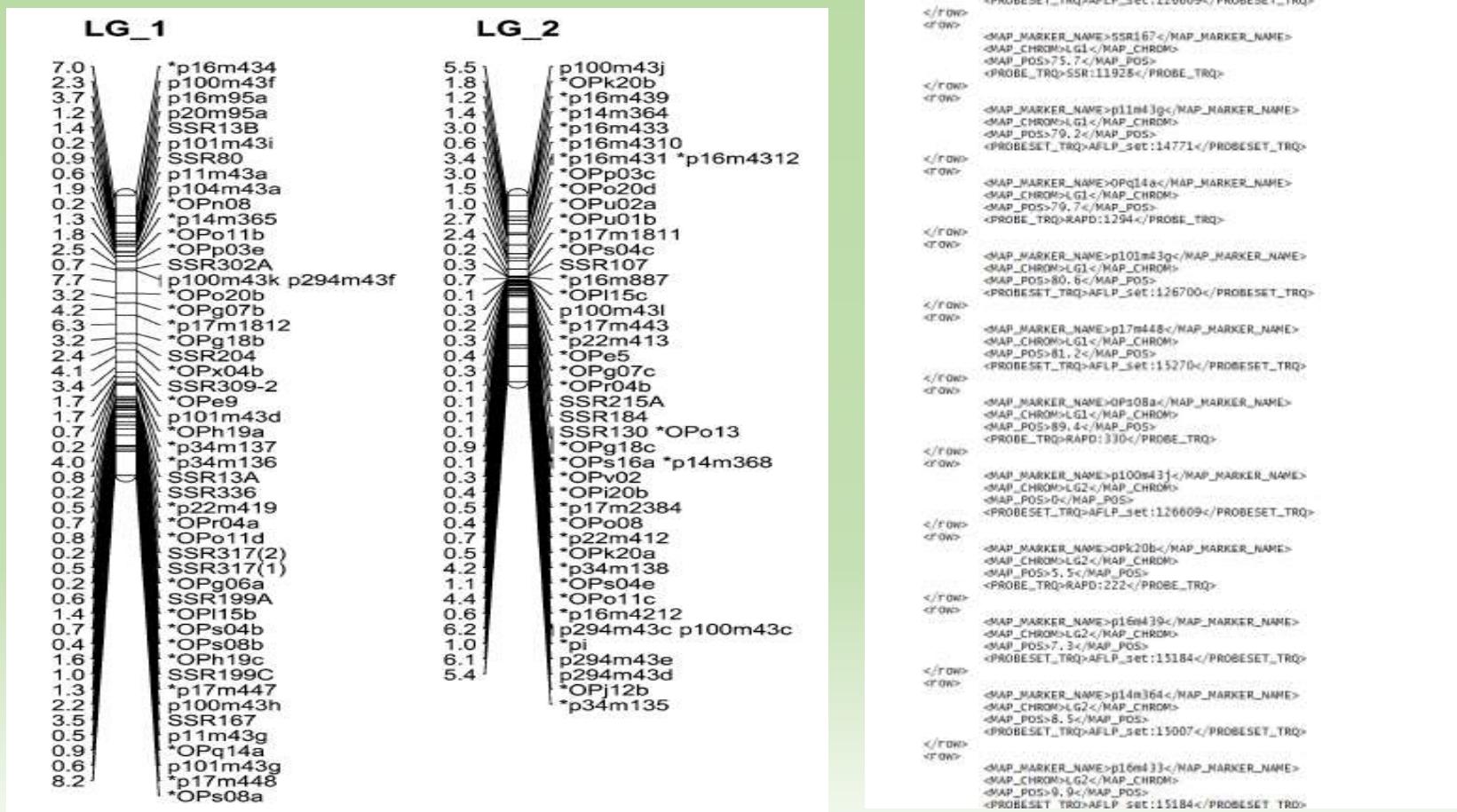
Genetic map processing



Genetic map processing

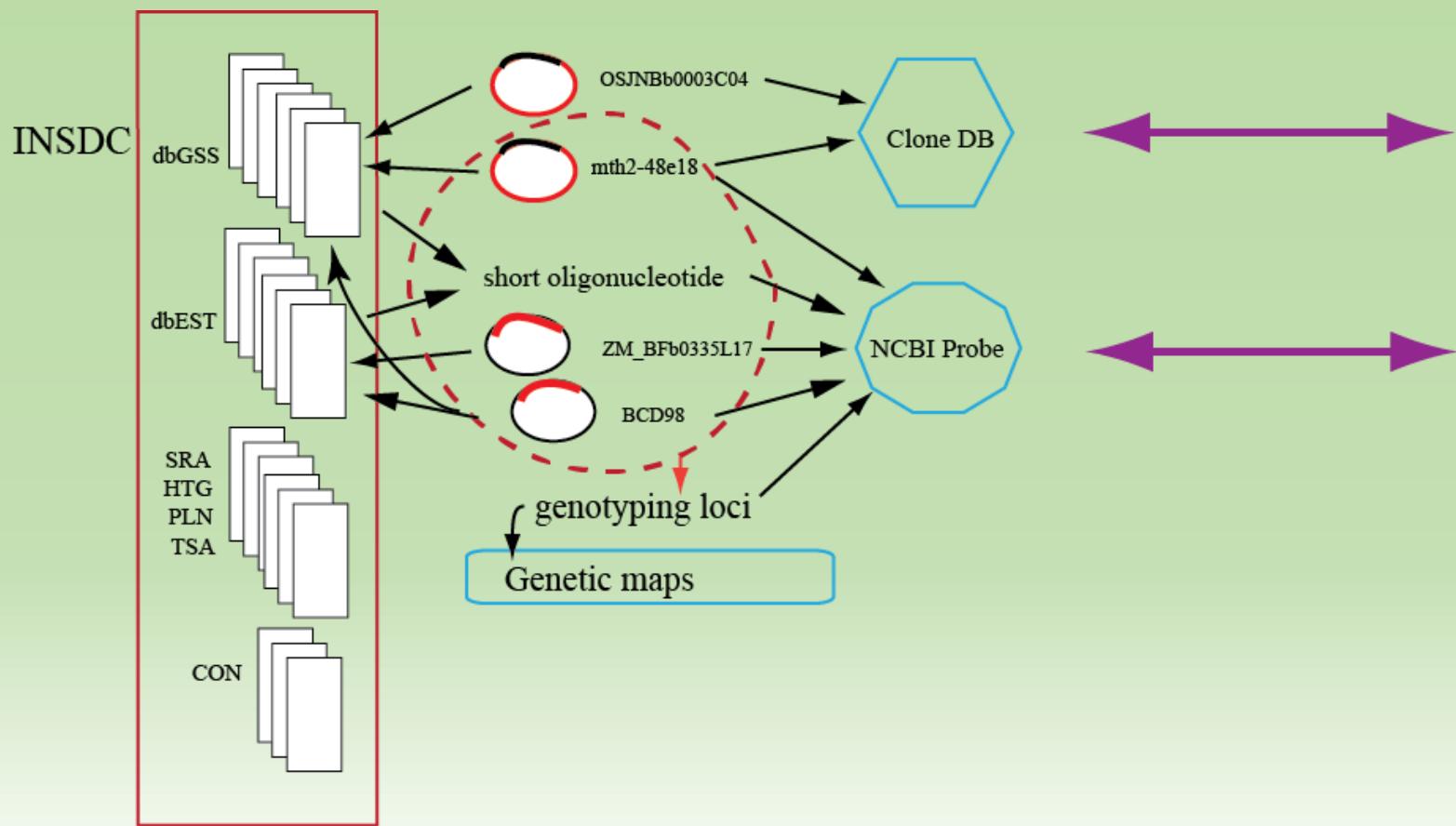


Genetic map processing

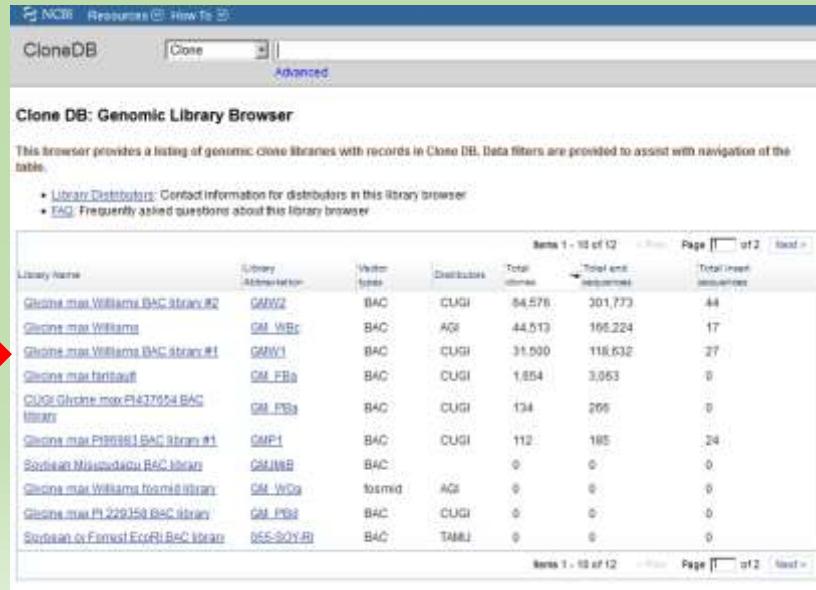


CloneDB -

<http://www.ncbi.nlm.nih.gov/clone/>



CloneDB – metadata and placement



The screenshot shows a web-based library browser. At the top, there's a navigation bar with links for 'NCBI Resources', 'How To', and 'CloneDB'. Below that is a search bar with 'Clone' and 'Advanced' buttons. The main title is 'Clone DB: Genomic Library Browser'. A sub-instruction says: 'This browser provides a listing of genomic clone libraries with records in Clone DB. Data filters are provided to assist with navigation of the table.' Below this are two links: 'Library Distributors' (Contact information for distributors in this library browser) and 'FAQ' (Frequently asked questions about this library browser). The main content is a table with the following data:

| Library Name | Library Administration | Vector type | Distributors | Total clones | Total cDNA sequences | Total insert sequences |
|--------------------------------------|------------------------|-------------|--------------|--------------|----------------------|------------------------|
| Clone max Williams BAC library #2 | GM_WB2 | BAC | CUGI | 54,576 | 301,773 | 44 |
| Clone max Williams | GM_WB | BAC | AGI | 44,513 | 165,224 | 17 |
| Clone max Williams BAC library #1 | GM_WB1 | BAC | CUGI | 31,500 | 116,632 | 27 |
| Clone max Harvard | GM_FBa | BAC | CUGI | 1,654 | 3,663 | 0 |
| CUGI Ghrote max PI437654 BAC library | GM_PSA | BAC | CUGI | 134 | 266 | 0 |
| Clone max PI#881 BAC library #1 | GM_P1 | BAC | CUGI | 112 | 185 | 24 |
| Soybean MaxEntdb1 BAC library | GM_SEB | BAC | | 0 | 0 | 0 |
| Clone max Williams hybrid library | GM_WG | fosmid | AGI | 0 | 0 | 0 |
| Clone max PI 229154 BAC library | GM_PSA | BAC | CUGI | 0 | 0 | 0 |
| Soybean & Forest EcoRI BAC library | SES-SOY-R | BAC | TABU | 0 | 0 | 0 |

CloneDB – metadata and placement

Glycine max Williams BAC Library #1

Library Summary

| | |
|------------------------------------|-------------|
| Library Abbreviation: | GMW1 |
| Organism: | Glycine max |
| Vector type(s): | BAC |
| # clones Clone DB: | 31,600 |
| # insert sequences Clone DB: | 31,632 |
| # insert sequences Clone DB: | 27,796 |
| # clones with both ends sequenced: | 27,796 |

Library Details

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| | | | |
|-------------------------|------------------------|-----------------------|------------------------|
| Library segment: ALL | Cutter: Williams R2 | Sex: hermaphrodite | Tissue: embryo test |
|-------------------------|------------------------|-----------------------|------------------------|

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| Library segment | Vector Name | Vector Cloning Site(s) | Insert Cloning Site(s) | Vector Selection | Insert Proc. Method | Lab Host | Vector Delivery Method |
|-----------------|-------------|------------------------|------------------------|------------------|---------------------|----------|------------------------|
| 1 | pGEMBAC1 | HindIII | HindIII | chloramphenicol | Partial digestion | DH10B | electroporation |

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| | | | | |
|-----------------------|------------------------|-------------------|-----------------------------|----------------|
| Library segment: 1 | Avg Insert(kb): 150 | Coverage: 5.42 | Plate range(s): 1 to 105 | % Plated: 1 |
|-----------------------|------------------------|-------------------|-----------------------------|----------------|

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

Alternate Library name(s):
ISU HindIII BAC Library

Alternate Library abbreviation(s):
GM_WBA, Williams R2, GM_WB01

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| | | | |
|---|-------------------------------|--|----------------------|
| Distributor Logo:  | Distributor Type: Academic | Contact: Clemson University Genomics Institute (CUGI) Email: orders@genome.clemson.edu | More Info Address |
|---|-------------------------------|--|----------------------|

CloneDB – metadata and placement

Glycine max Williams BAC Library #1

Library Summary

| | |
|------------------------------------|-------------|
| Library Abbreviation: | GMW1 |
| Organism: | Glycine max |
| Vector (vector): | BAC |
| # clones Clone DB: | 31,600 |
| # end sequences Clone DB: | 116,832 |
| # insert sequences Clone DB: | 27 |
| # clones with both ends sequenced: | 27,796 |

LIBRARY DETAILS

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| | | | |
|------------------|-----------------------|--------------------|------------------------|
| Library segment: | Cultivar: Williams 82 | Sex: hermaphrodite | Tissue: etiolated leaf |
| ALL | | | |

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| | | | | | | | |
|------------------|--------------------------|---------------------------------|---------------------------------|--|---|-----------------|---|
| Library segment: | Vector Name: pGEM-BAC(1) | Vector Cloning Strain(s): JM101 | Insert Cloning Strain(s): JM101 | Vector Selection Method: chlorampricin | Insert Prep Method: enzymatic digestion | Lac Host: DH10B | Vector Delivery Method: electroporation |
| 1 | | | | | | | |

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| | | | | |
|------------------|---------------------|----------------|--------------------------|-------------|
| Library segment: | Avg Insert(kb): 150 | Coverage: 5.42 | Plate range(s): 1 to 105 | % Plated: 1 |
| 1 | | | | |

DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| | |
|----------------------------|------------------------------------|
| Alternate library name(s): | Alternate library abbreviation(s): |
| ISU Hender BAC Library | GM_W82A, W82A_B2, GM_W8001 |

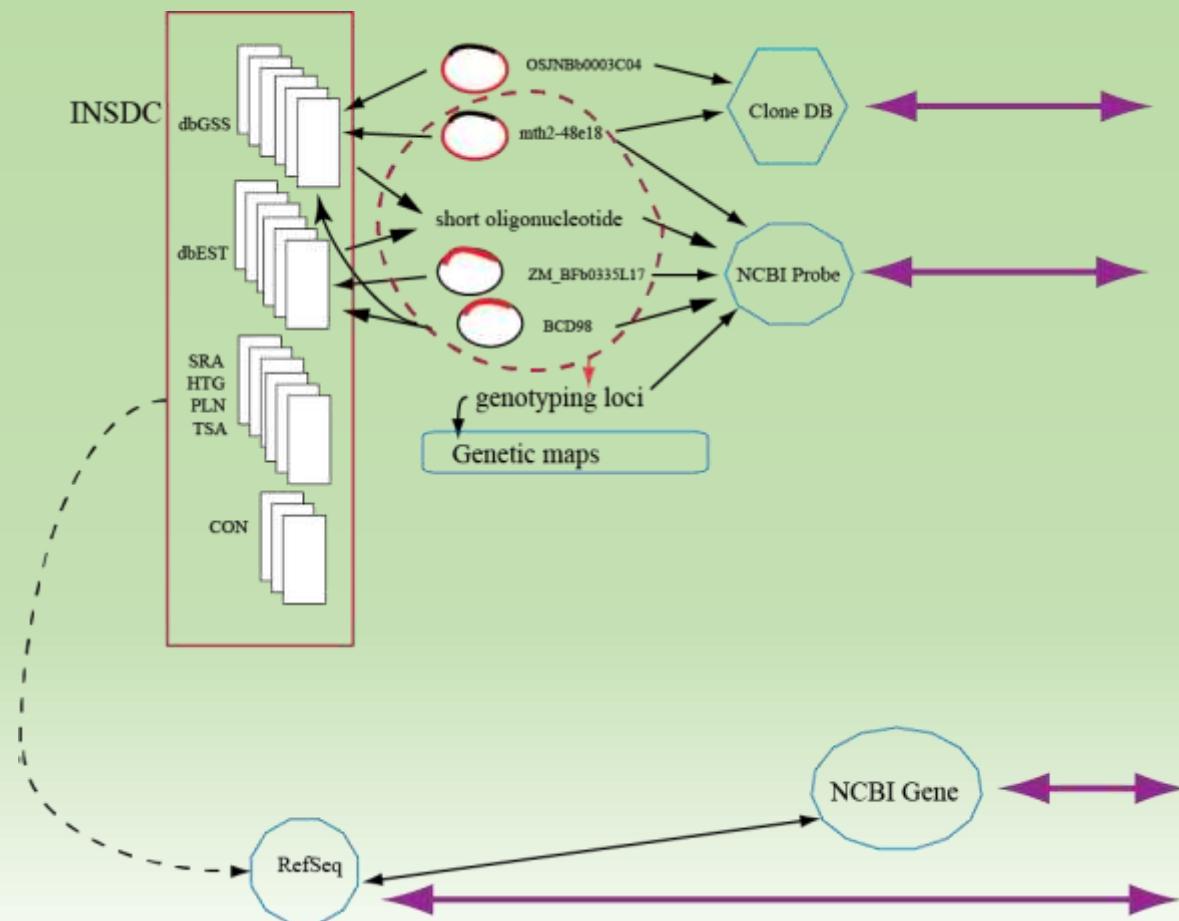
DNA Source | Library Construction | Library Statistics | Alternate Library Names | Distributors

| | | | |
|------------------------|------------------------------------|---|---------------------|
| Distributor Logo: CUGI | Distributor Type: Creator Academic | Contact: Clemson University Genomics Institute (CUGI) Email: orders@genome.clemson.edu | More Info Address |
|------------------------|------------------------------------|---|---------------------|



NCBI RefSeq -

<http://www.ncbi.nlm.nih.gov/refseq/>



Single source, multiple sources

Nucleotide Nucleotide Advanced

Display Settings: GenBank

Send:

Zea mays CASP-like protein 13 (LOC100276350), mRNA

NCBI Reference Sequence: NM_001150159.1

FASTA Graphics

Gene:

Locus: NM_001150159 872 bp mRNA linear PLN 12-OCT-2014

Definition: Zea mays CASP-like protein 13 (LOC100276350), mRNA.

Accession: NM_001150159

Version: NM_001150159.1 GI:226500149

Keywords: RefSeq.

Source: Zea mays

Organism: Zea mays
Eukaryota; Viridiplantae; Streptophytina; Embryophytina; Tracheophytina;
Spermatophytina; Magnoliophytina; Liliopsida; Poales; Poaceae; Pooideae
clade; Panicoideae; Andropogoneae; Zea.

Reference: 1 (bases 1 to 872)

Authors: Alexandrov SB, Brover VV, Freidin S, Troukhman NF, Tatarinova TV,
Zhang H, Swaller TJ, Lu YF, Bouck J, Flavell RB and Feldmann RA.

Title: Insights into corn genes derived from large-scale cDNA sequencing

Journal: Plant Mol Biol 70: 177-190 (2006) [PubMed]

Published: 18937034

Comment: PROVISIONAL REVIEW: This record has not yet been subject to final NCBI review. The reference sequence was derived from EU961555.1.
##Evidence-Data-START##
Transcript exon combination :: EU961555.1 [ECO:0000332]
##Evidence-Data-END##

Nucleotide Nucleotide Advanced

Display Settings: GenBank

Send:

Homo sapiens toll-like receptor 4 (TLR4), transcript variant 3, mRNA

NCBI Reference Sequence: NM_003266.3

FASTA Graphics

Gene:

Locus: NM_003266 5751 bp mRNA linear PRI 25-MAY-2014

Definition: Homo sapiens toll-like receptor 4 (TLR4), transcript variant 3, mRNA.

Accession: NM_003266 NR_024168

Version: NM_003266.3 GI:373432601

Reinberg: RefSeq.

Source: Homo sapiens (human)

Organism: Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Hominoidea;
Catarrhini; Hominidae; Homo;

Reference: 1 (bases 1 to 5751)

Authors: Chen YF, Jhau AL, Wang SH, Yang JS, Chang SJ, Wu CL and Wu JS.

Title: Shankui acid A isolated from Taiwanofungus camphoratus is a novel selective TLR4/MD-2 antagonist with anti-inflammatory properties

Journal: J. Immunol. 192 (6): 2778-2786 (2014)

Published: 9435238

Comment: Rock JL, Hardiman G, Timans JC, Kastlein EA and Baran JF.
A family of human receptors structurally related to Drosophila Toll.
Proc. Natl. Acad. Sci. U.S.A. 95 (2): 588-593 (1998)
Medzhitov R, Preston-Hurlburt P and Janeway CA Jr.
A human homologue of the Drosophila Toll protein signals activation of adaptive immunity
Nature 388 (6647): 334-337 (1997)
9237759
REVIEWED REVIEW: This record has been curated by NCBI staff. The reference sequence was derived from DA187988.1, BC149841.1,
BC025294.1, AK216152.1, BG622357.1 and DR218506.1.
On or before Jan 21, 2012 this sequence version replaced
gi:207028451, gi:19924147.

Toll-like receptor (TLR) family which plays a fundamental role in pathogen recognition and activation of innate immunity. TLRs are highly conserved from Drosophila to humans and share structural and functional similarities. They recognize pathogen-associated molecular patterns that are expressed on infectious agents, and

Single source, multiple sources

```

transcript exon combination :: eug1555.1 [ECO:0000352]
##Evidence-Data-END##
FEATURES
source      Location/Qualifiers
              1..872
              /organism="Zea mays"
              /mol_type="mRNA"
              /db_xref="taxon:4577"
              /chromosome="7"

gene        1..872
              /gene="LOC100276350"
              /note="CASP-like protein 13"
              /db_xref="GeneID:100276350"
              ..._etc

              /gene="LOC100276350"
              /note="zmCASPL5B3"
              /codon_start=1
              /product="CASP-like protein 5B3"
              /protein_id="NP_001143631.1"
              /db_xref="GI:226508150"
              /db_xref="GeneID:100276350"
              /translation="MKDWWGSPGIWGSMAIRLSQCVSAGASNGAMATAYGFSNTTAFCA
YIILASMGQLQLMSFGCLCDVYVSLKTRKRLHNPKVVLVSLEVVGDWVIALSFAAASASA
GVVILIFERDWFHFCRMPQLSCGRYALSVLAFITWSFIATSAVSMEWLLPSL"
misc_feature 215..283
              /gene="LOC100276350"
              /experiment="experimental evidence, no additional details
recorded"
              /note="propagated from UniProtKB/Swiss-Prot (B6I990.1);
transmembrane region"
              ..._etc
misc_feature 299..361
              /gene="LOC100276350"
              /experiment="experimental evidence, no additional details
..._etc

```

ERS025082, ERS025084 [ECO:0000348]

```

##Evidence-Data-END##
COMPLETENESS: complete on the 3' end.
PRIMARY REFSEQ_SPAN PRIMARY_IDENTIFIER PRIMARY_SPAN COMP
1-278  DA187988.1 1-278
279-2980 BC143841.1 1-2702
2981-3242 BC025294.1 2480-2741
3243-5161 AR226155.1 2998-4916
5162-5608 BG622357.1 121-567
5609-5781 DA218006.1 386-558

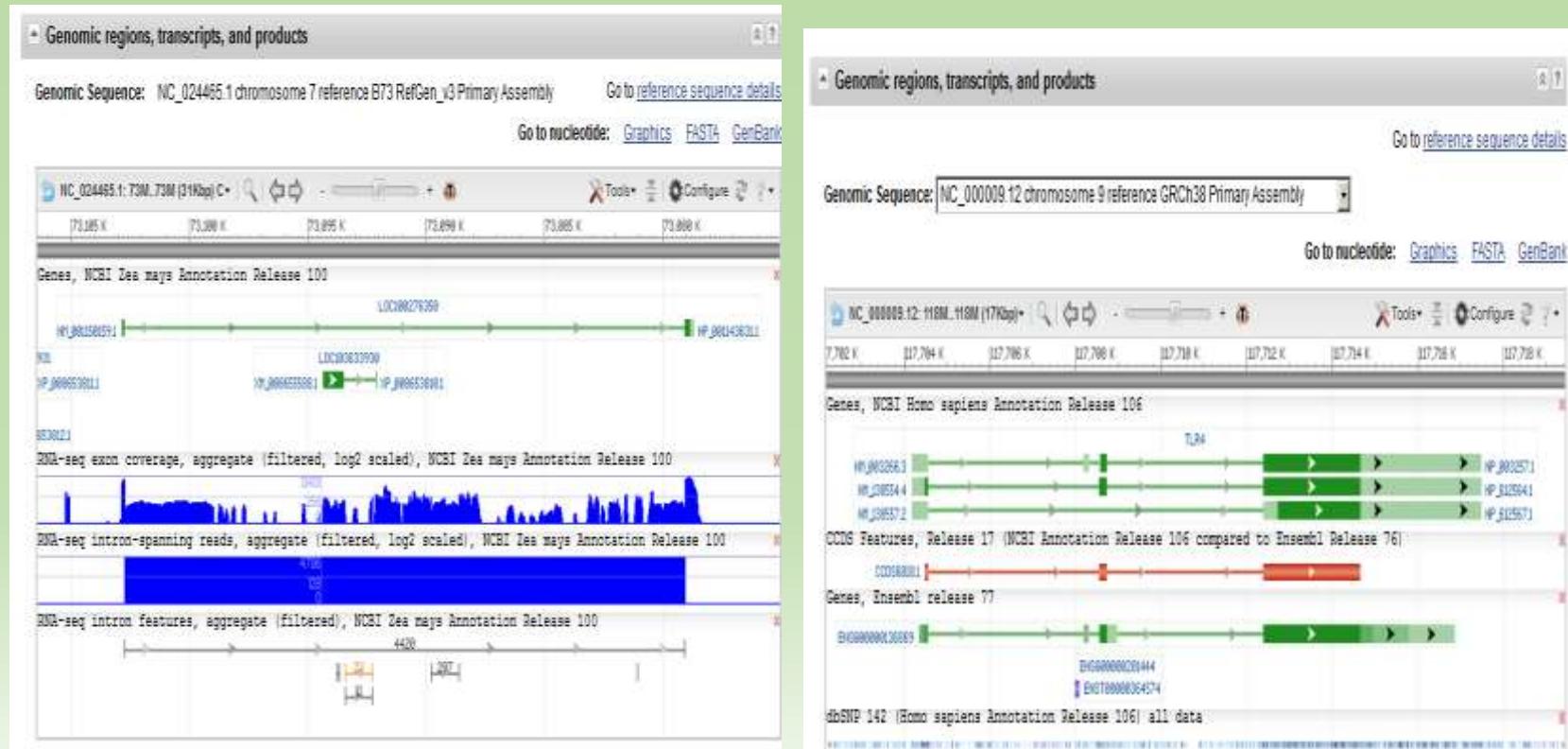
FEATURES
source      Location/Qualifiers
              1..5781
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /chromosome="3"
              /map="3q33.1"

gene        1..5781
              /gene="ILR4"
              /gene_synonym="ARMND10; CD284; TLR-4; TOLL"
              /note="toll-like receptor 4"
              /db_xref="GeneID:7099"
              /db_xref="HGNC:HGNC:11250"
              /db_xref="MIM:603030"
              ..._etc

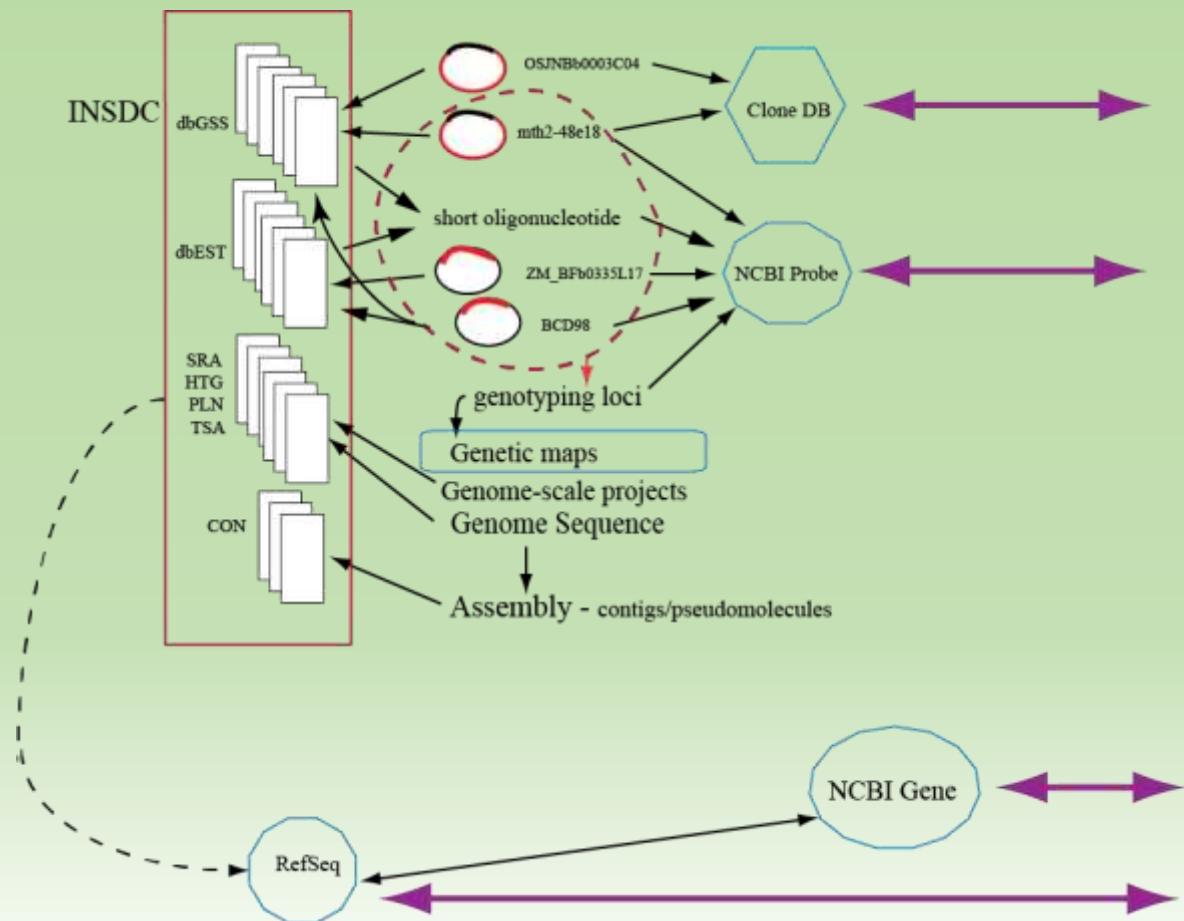
exon       1..5781
              /gene="ILR4"
              /gene_synonym="ARMND10; CD284; TLR-4; TOLL"
              /inference="alignment:Splign:1.39.8"
              392..511
              /gene="ILR4"
              /gene_synonym="ARMND10; CD284; TLR-4; TOLL"
              /inference="alignment:Splign:1.39.8"
              misc_feature 401..403
              /gene="ILR4"
              /gene_synonym="ARMND10; CD284; TLR-4; TOLL"
              /note="upstream in-frame stop codon"
              512..678
              /gene="ILR4"
              /gene_synonym="ARMND10; CD284; TLR-4; TOLL"
              molecular patterns that are expressed on infectious agents, and

```

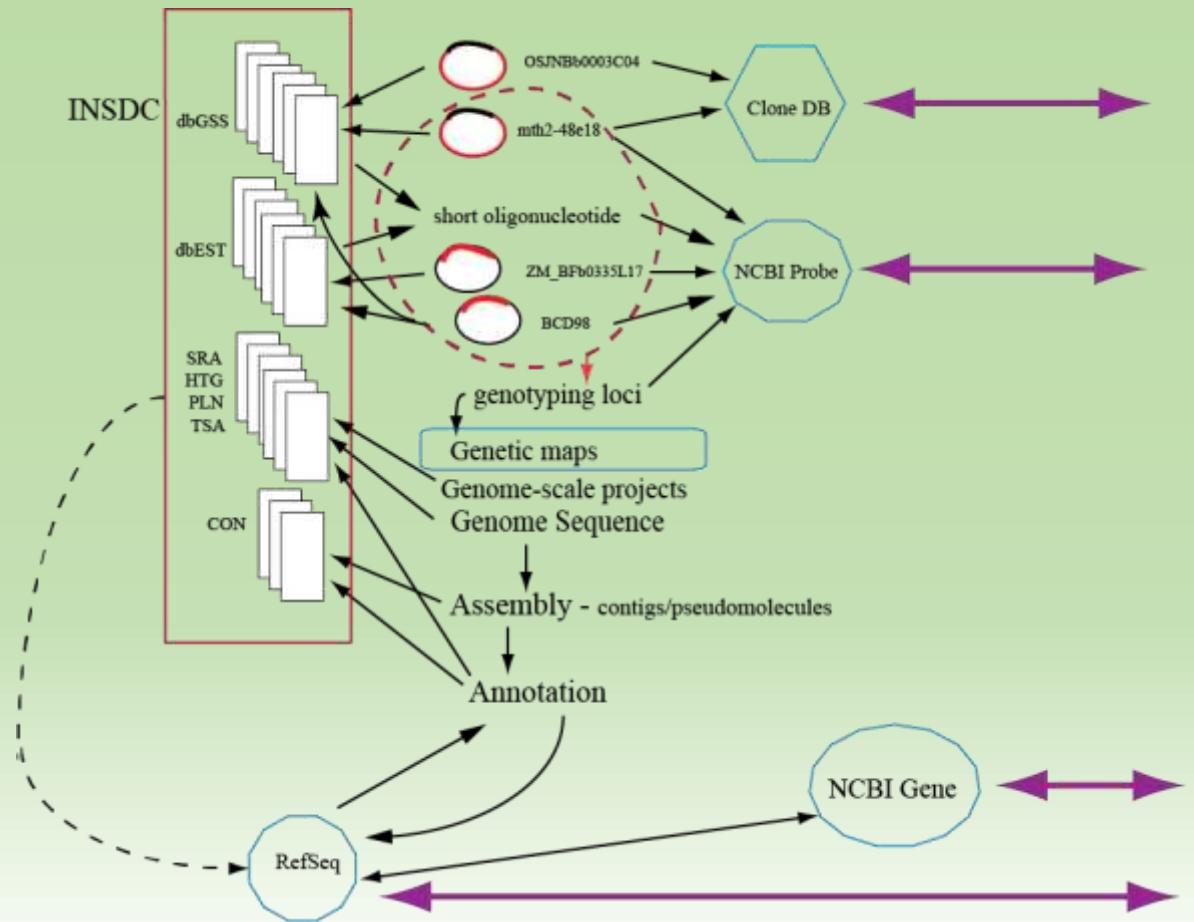
Gene to RefSeq one-to-one or one-to-many



Genome sequence



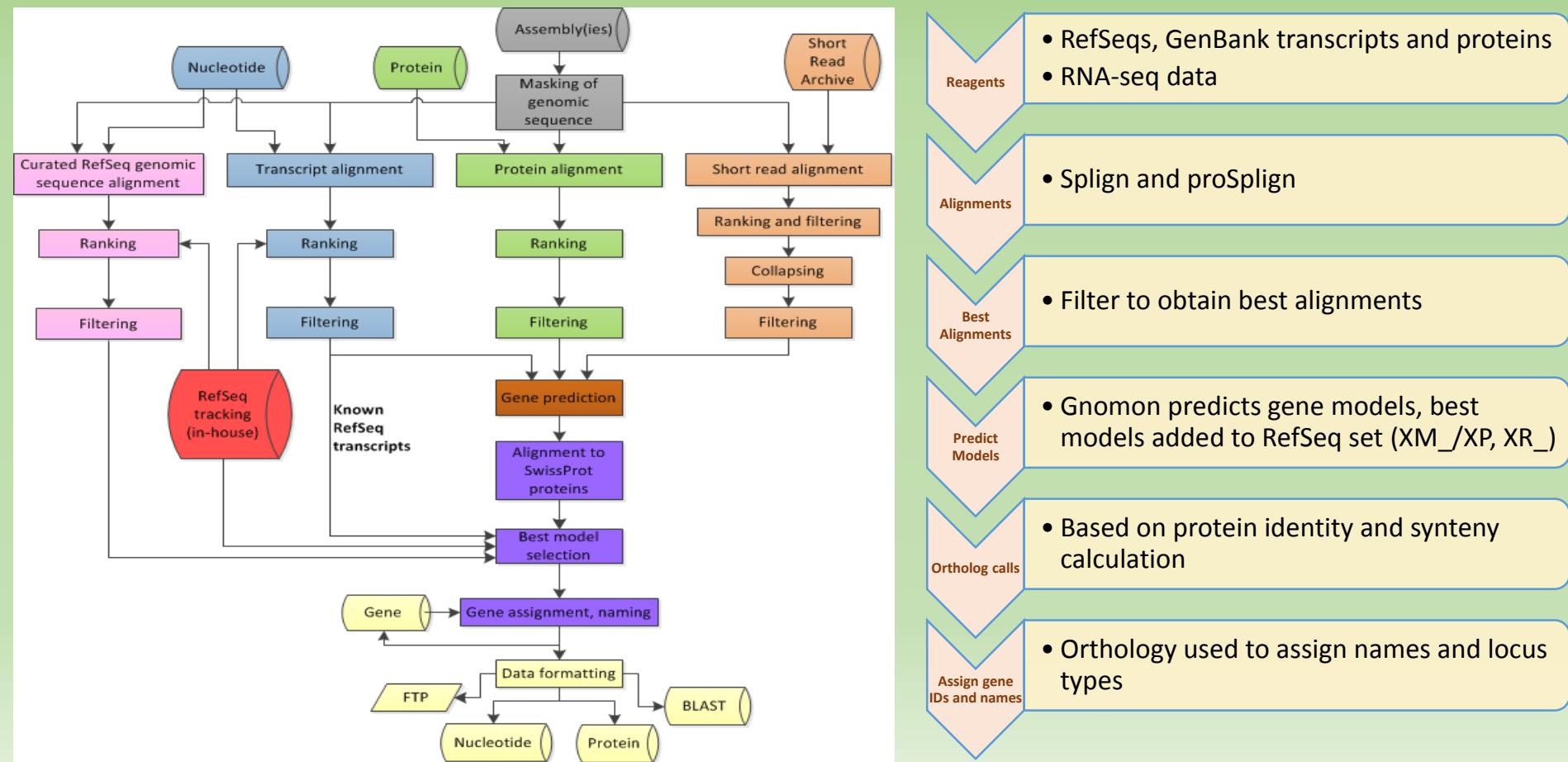
Annotation



Eukaryotic Genome Annotation

- Input “substrate” is a RefSeq genome (contigs or chromosomes)
- Input “reagents” are sequence records:
 - INSDC mRNA
 - INSDC protein
 - EST
 - RNA-seq
 - NM/NP RefSeq
- Output is a collection of RefSeq nuc/prot pairs, each pair belonging to an associated NCBI Gene record

Eukaryotic Genome Annotation



*Speed (5-10 days), *RNA-seq data for gene prediction, *using aligned transcripts for 'gap-filling', *enhanced public reports

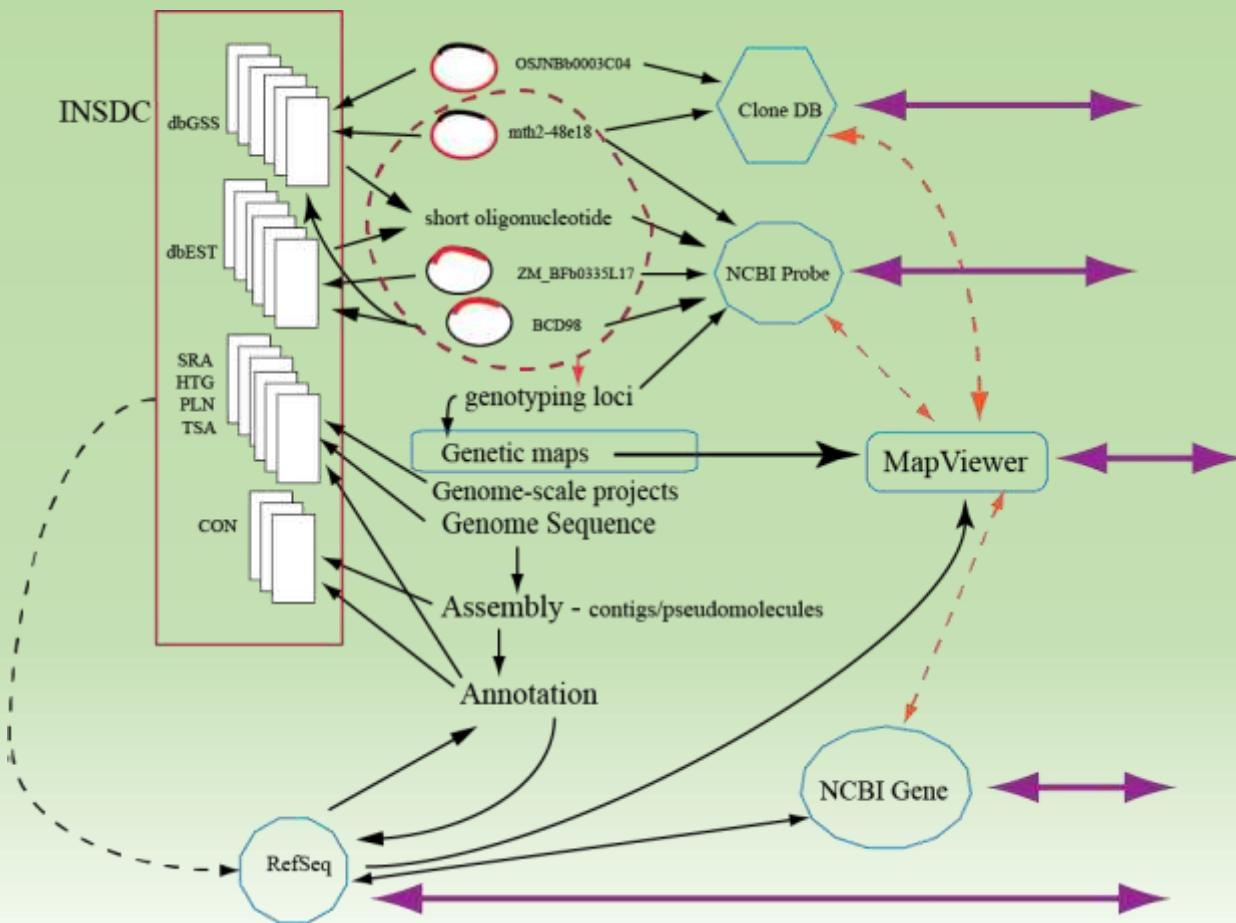
Visit Poster P0051 for more on genome annotation

Plant genomes – as of 2015

- 146 plants have 246 nuclear genomes submitted
 - 43 of these plants have chromosome assemblies
- 561 plants have only organelle genomes submitted
- 31 plants have annotated genomes
 - 4 are presented with the submission annotation

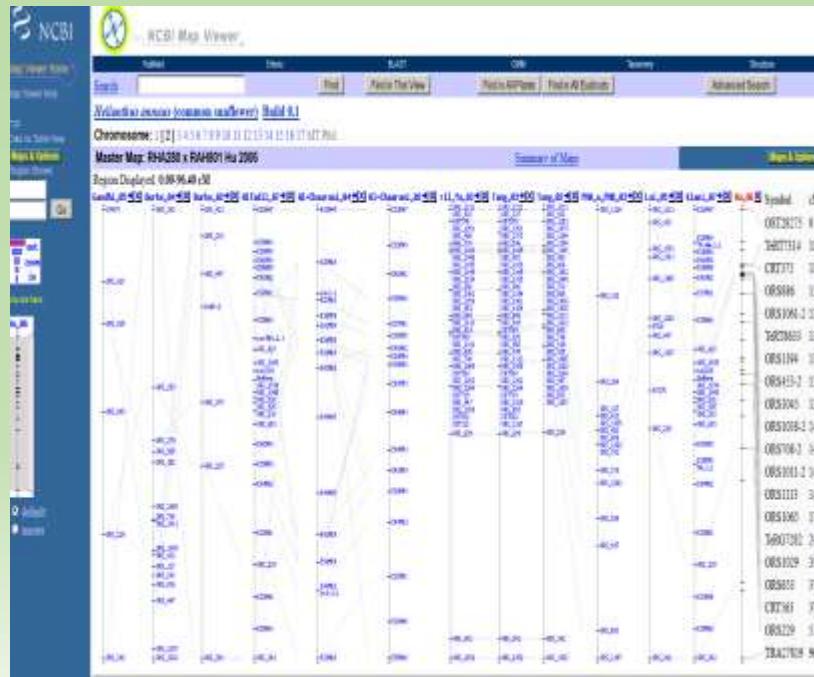
Map Viewer -

<http://www.ncbi.nlm.nih.gov/mapview/>

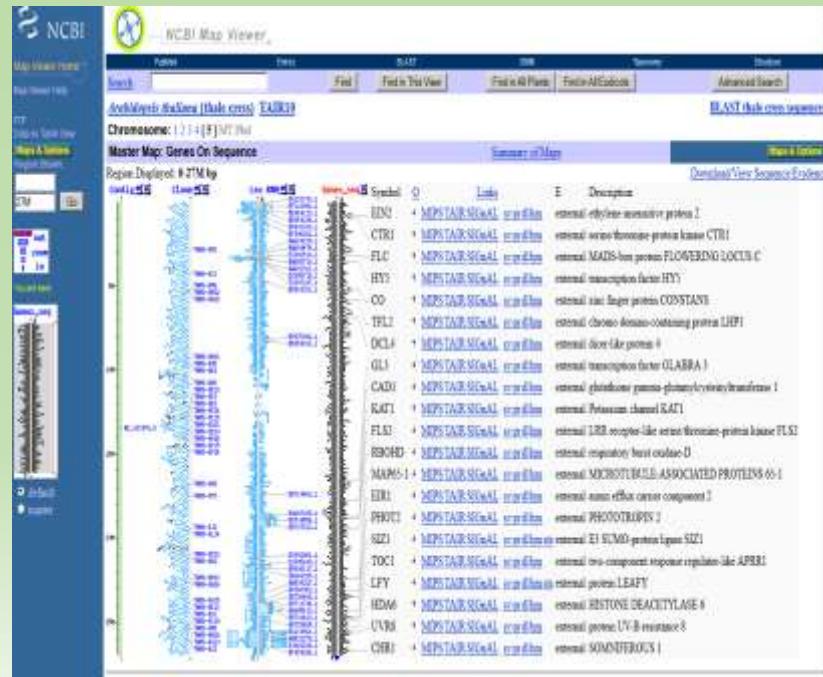


Map Viewer – display tool for genome units

Genetic Maps



Sequence Maps



Text search for gene name – the backdoor to unplaced sequence

The screenshot shows the NCBI Map Viewer interface for *Cucumis melo*. A red box highlights the search bar at the top, which contains the text "on the chromosome". Below the search bar, the results page displays a message stating that no scaffolds or contigs have been placed on chromosomes, and provides instructions for finding sequence data by querying with a gene symbol or accession number. It also lists two methods for identifying genes of interest: searching the Gene database or performing a BLAST search against the genome. At the bottom, there are two diagrams showing the linkage of genes to chromosomes, labeled "Lineage" and "Cucumis melo".

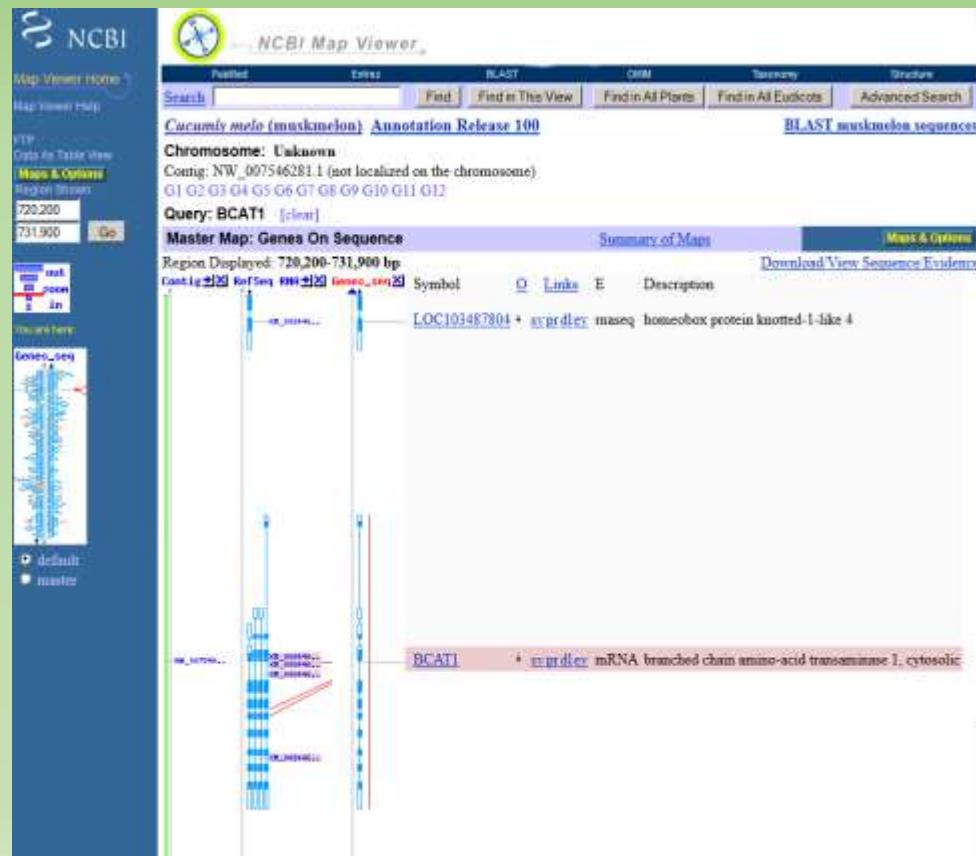
Text search for gene name – the backdoor to unplaced sequence

The figure consists of two side-by-side screenshots of the NCBI Map Viewer. Both screenshots show the search results for the gene 'BCAT1'.

Left Screenshot: The search bar at the top contains 'on the chromosome(s)' and 'assembly All'. A red box highlights the search bar. Below the search bar, the page title is 'Cucumis melo (muskmelon) genome view' and it says 'Annotation Release 100 statistics'. A yellow box highlights the text: 'At present, none of the scaffolds or contigs from the current public WGS assembly have been placed on chromosomes so you cannot browse the sequence data by chromosome. You can find sequenced regions of interest by querying Map Viewer with a gene symbol or accession number. You can also identify genes of interest by:'. Below this text is a bulleted list: '• Searching [Gene](#) for muskmelon genes and following the link to Map Viewer.' and '• Searching the sequence using [BLAST](#) against the genome. Results from a BLAST sequence alignment can be displayed in Map Viewer by following the link in 'Related Information' to the right of the alignment.' At the bottom, there are two diagrams showing genomic tracks for chromosomes 91, 92, 93, 94, 95, 96, 97, and 98. The tracks are mostly black with some blue segments. A legend at the bottom left indicates: Lineage: Eukaryota; Viridiplantae; Streptophytina; Embryophytina; Tracheophytina; Spermatophytina; Mammophytina; eudicots; lemons; Cucurbitaceae; Pentapetalous rosids; Fabids; Cucurbitaceae; Cucurbitaceae; Benincasaceae; Cucurbitaceae; Cucumis melo.

Right Screenshot: The search bar at the top contains 'BCAT1'. Below the search bar, the page title is 'Cucumis melo (muskmelon) genome view' and it says 'Annotation Release 100 statistics'. A yellow box highlights the text: 'Expand by the items'. Below this is a diagram showing genomic tracks for chromosomes 91, 92, 93, 94, 95, 96, 97, and 98. The tracks are mostly black with some blue segments. A legend at the bottom left indicates: Lineage: Eukaryota; Viridiplantae; Streptophytina; Embryophytina; Tracheophytina; Spermatophytina; Mammophytina; eudicots; lemons; Cucurbitaceae; Pentapetalous rosids; Fabids; Cucurbitaceae; Cucurbitaceae; Benincasaceae; Cucurbitaceae; Cucumis melo. Below the tracks is a table titled 'Search results for query "BCAT1": 6 hits'. The table has columns: Chr, Mech, Map element, Type, and Mag. The data is as follows:

| Chr | Mech | Map element | Type | Mag |
|----------|--|--------------|------------|------------|
| Unplaced | Cucumis melo branched-chain BCAT1 mRNA, partial cds | FT008171 | TRANSCRIPT | RefSeq RNA |
| Unplaced | PREDICTED: Cucumis melo branched-chain-amino-acid aminotransferase | XM_004462881 | TRANSCRIPT | RefSeq RNA |
| Unplaced | PREDICTED: Cucumis melo branched-chain-amino-acid aminotransferase | XM_004462881 | TRANSCRIPT | RefSeq RNA |
| Unplaced | PREDICTED: Cucumis melo branched-chain-amino-acid aminotransferase | XM_004462881 | TRANSCRIPT | RefSeq RNA |
| Unplaced | PREDICTED: Cucumis melo branched-chain-amino-acid aminotransferase | XM_004462881 | TRANSCRIPT | RefSeq RNA |
| Unplaced | BCAT1 - branched-chain amino-acid transaminase 1, cytosolic | BCAT1 | GENE | Gene seq |



BCAT1 branched chain amino-acid transaminase 1, cytosolic [Cucumis melo (muskmelon)]

Gene ID: 103487805, updated on 9-Aug-2014

Summary

Gene symbol: BCAT1
Gene description: branched chain amino-acid transaminase 1, cytosolic
Gene type: protein coding
RefSeq status: MODEL
Organism: Cucumis melo
Lineage: Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytina; Magnoliophytina; eudicots; Rosids; Asterids; Rosidae; Cucurbitales; Cucurbitaceae; Benincasaceae; Cucumis

Genomic context

Location: chromosome: Un
Exon count: 11

| Annotation release | Status | Assembly | Chr | Location |
|--------------------|---------|-------------------------------|-------------------|------------------------------------|
| 100 | current | ASM3130v1 (GCF_00031304v1) | Unplaced Scaffold | NW_007546281.1 (723696..728371) |

NW_007546281.1

Genomic regions, transcripts, and products

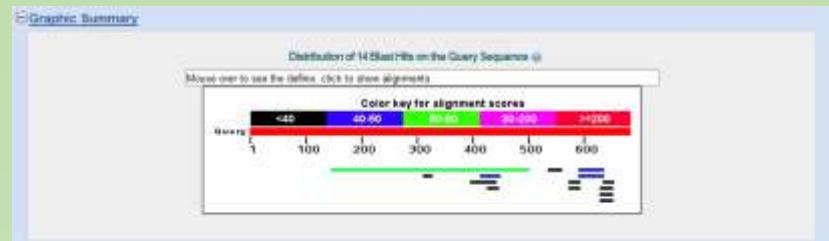
Genomic Sequence: NW_007546281 unplaced scaffold reference ASM3130v1 Primary Assembly/maseq sequence details



Finding an ortholog and the genome context

- Fruit weight QTL identified in tomato in 1990s
- The tomato gene was cloned in 2000
- Are there fruit weight orthologs in other crops such as apple, melon, or grape?

BLAST of melon RefSeq mRNA with tomato mRNA



Graphic Summary

Show Conserved Domains

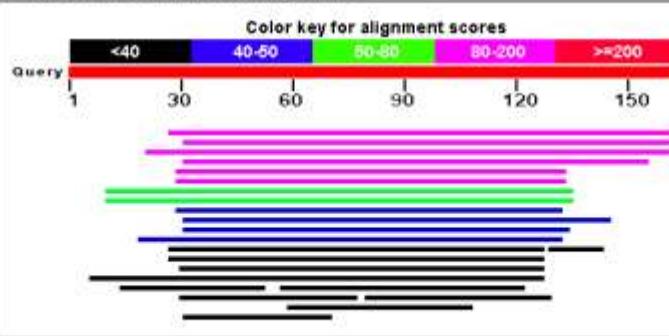
Putative conserved domains have been detected, click on the image below for detailed results.

Query seq.
Specific hits
Superfamilies



Distribution of 23 Blast Hits on the Query Sequence

Mouse over to see the details, click to show alignments



Graphic Summary

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

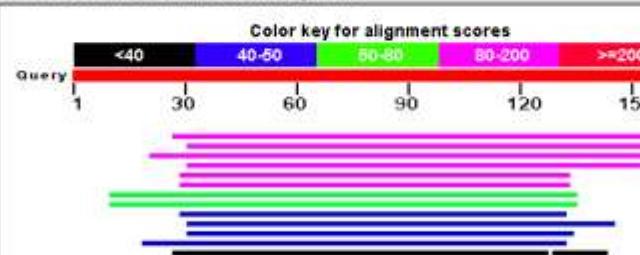
Query seq.
Specific hits
Superfamilies

Pfam

PLACB superfamily

Distribution of 23 Blast Hits on the Query Sequence ⓘ

Mouse over to see the details, click to show alignments



Download ▾ GenPept Graphics

▼ Next ▲ Previous ▲ Descriptions

PREDICTED: cell number regulator 2-like [Cucumis melo]

Sequence ID: ref|XP_008464696.1| Length: 150 Number of Matches: 1

Range 1: 17 to 149 GenPept Graphics

▼ Next Match ▲ Previous Match

| Score | Expect | Method | Identities | Positives | Gaps |
|---------------|--------|------------------------------|-------------|-------------|-----------|
| 154 bits(388) | 3e-47 | Compositional matrix adjust. | 69/133(52%) | 95/133(71%) | 1/133(0%) |

Query 32 WSTGLCHCFDDPANCLVTSVCPCTIFGQISEIILNKGTTS CGSRGALYCLLG-LTGLPSLY 90
WS+GLC CF D +C T+ CPC+ FGQ SEI+++G-TSC ++CL+ T LY
Sbjct 17 WSSGLCDCFSMDMGSCCCITWCPCVPFGQASEIIIDEGSTISCFGNALIFCLIASFTPCICLY 76

Query 91 SCFYRSKMRGQYDLEEAPCVDCLVHVFC EPCALCQEYRELKNRGFDMGIGWQANMDRQSR 150
+C YRS++R +Y+L+E PC DC VH +C CA+CQEYRELKNRGRF+M IGWQ N+ RQ++
Sbjct 77 TCSYRSRLRKYNLKEPCNDCCVHCWCSCAMCQEYRELKNRGRFNMHIGWQENVQRQNK 136

Query 151 GVTMPPYHAGMTR 163
G+ +PP G +
Sbjct 137 GIEIPPTVPGQMK 149

Related Information

[Gene](#) - associated gene details

[Display Settings](#): Full Report

[Send to:](#)

 Filters activated: Current only. [Clear all](#) to show 1 items.

LOC103502518 cell number regulator 2-like [*Cucumis melo* (muskmelon)]

Gene ID: 103502518, updated on 9-Aug-2014

Summary

Gene symbol: LOC103502518
Gene description: cell number regulator 2-like
Gene type: protein coding
RefSeq status: MODEL
Organism: *Cucumis melo*
Lineage: Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Gunneridae; Pentapetalae; rosids; fabids; Cucurbitales; Cucurbitaceae; Benincaseae; Cucumis

Genomic context

Location: chromosome: Un

Exon count: 4

| Annotation release | Status | Assembly | Chr. | Location |
|--------------------|---------|------------------------------|-------------------|---|
| 100 | current | ASM31304v1 (GCF_000313045.1) | Unplaced Scaffold | NW_007546360.1 (52988..55967, complement) |

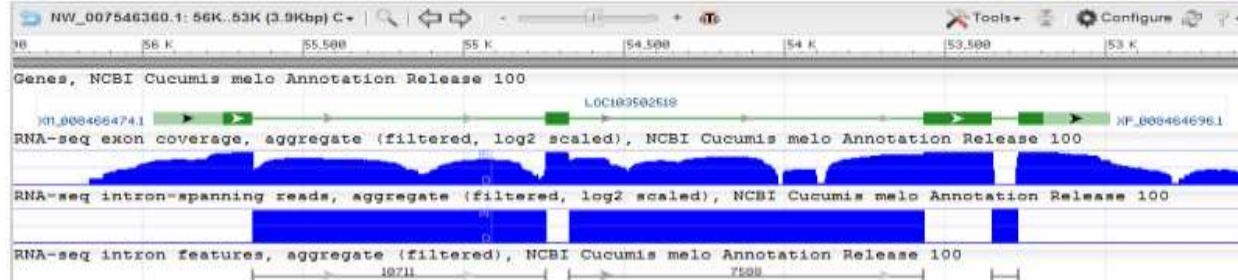


[▼ Next](#) [▲ Previous](#) [▲ Descriptions](#)

Genomic regions, transcripts, and products

Genomic Sequence: NW_007546360 unplaced scaffold reference ASM31304v1 Primary Assembly [Go to reference sequence details](#)

[Go to nucleotide:](#) [Graphics](#) [FASTA](#) [GenBank](#)



Related Information

[Gene](#) - associated gene details

Bibliography

[Display Settings](#): Full Report

[Send to:](#)

 Filters activated: Current only. [Clear all](#) to show 1 items.

LOC103502518 cell number regulator 2-like [*Cucumis melo* (muskmelon)]

Gene ID: 103502518, updated on 9-Aug-2014

Summary

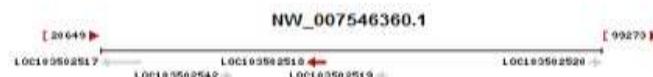
Gene symbol: LOC103502518
Gene description: cell number regulator 2-like
Gene type: protein coding
RefSeq status: MODEL
Organism: *Cucumis melo*
Lineage: Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Gunneridae; Pentapetalae; rosids; fabids; Cucurbitales; Cucurbitaceae; Benincaseae; Cucumis

Genomic context

Location: chromosome: Un

Exon count: 4

| Annotation release | Status | Assembly | Chr. | Location |
|--------------------|---------|------------------------------|-------------------|---|
| 100 | current | ASM31304v1 (GCF_000313045.1) | Unplaced Scaffold | NW_007546360.1 (52988..55987, complement) |

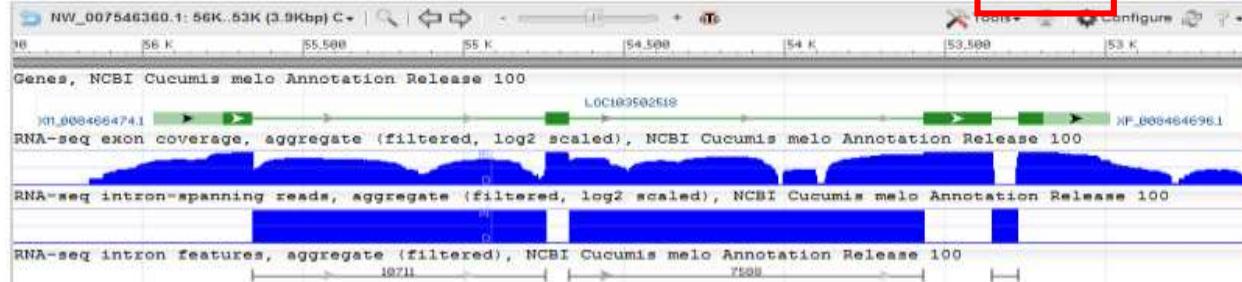


[▼ Next](#) [▲ Previous](#) [▲ Descriptions](#)

Genomic regions, transcripts, and products

Genomic Sequence: NW_007546360 unplaced scaffold reference ASM31304v1 Primary Assembly [Go to reference sequence details](#)

[Go to nucleotide](#) [Graphics](#) [FASTA](#) [GenBank](#)



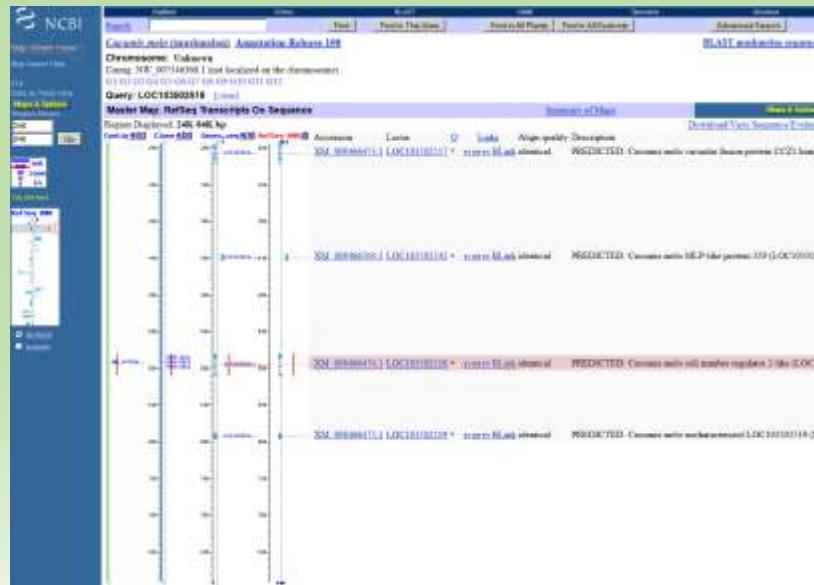
Related Information

[Gene](#) - associated gene details

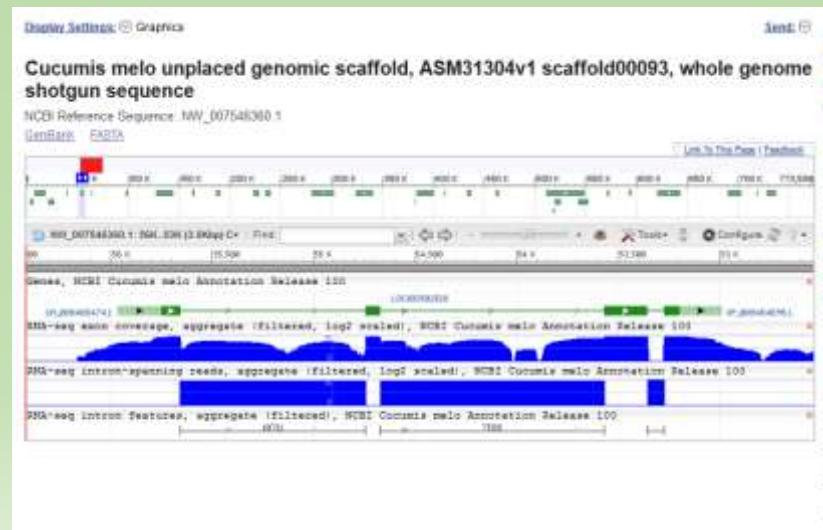
Bibliography

Finding ortholog(s)

Map Viewer by text search



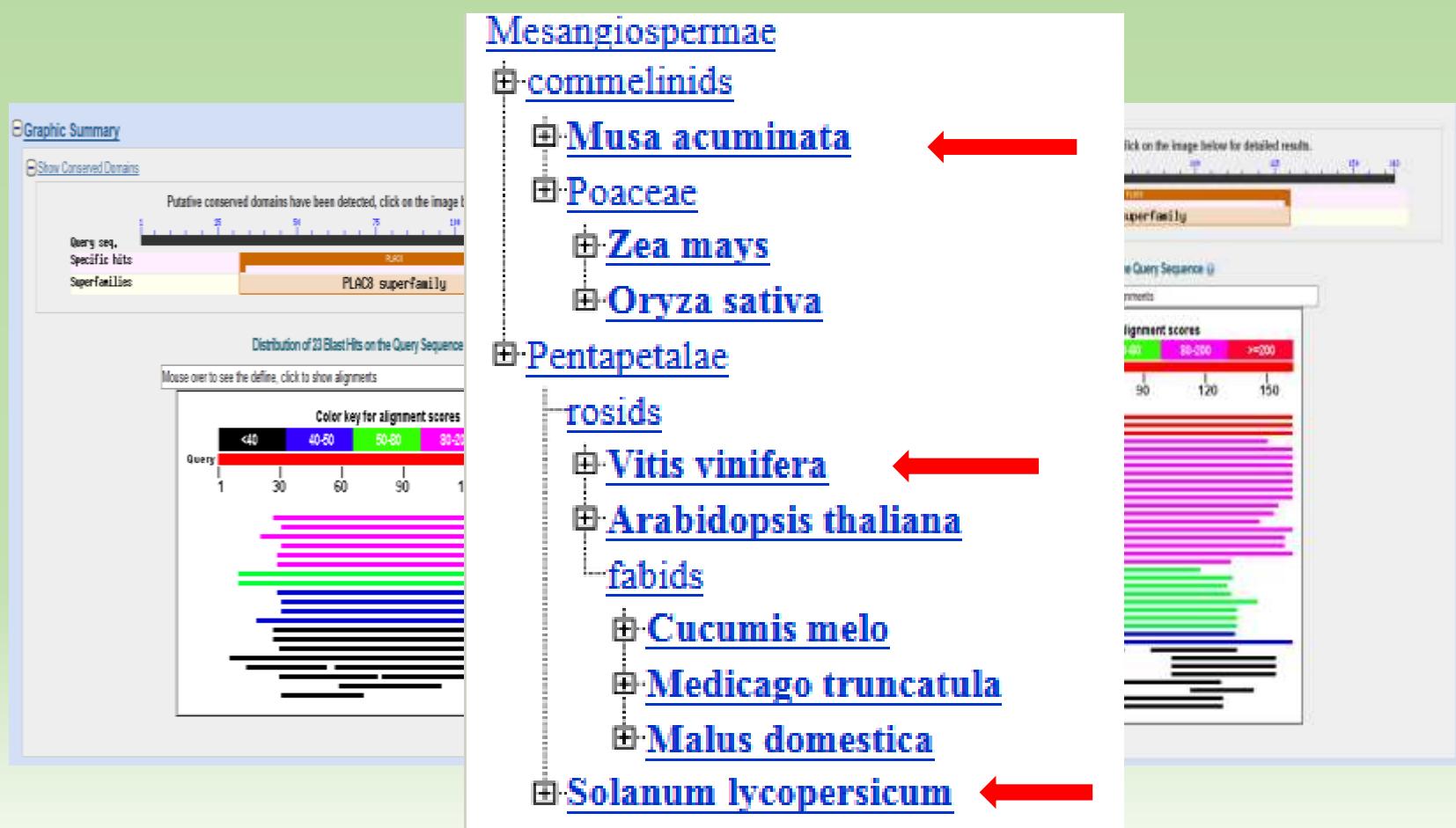
Sequence viewer by URL



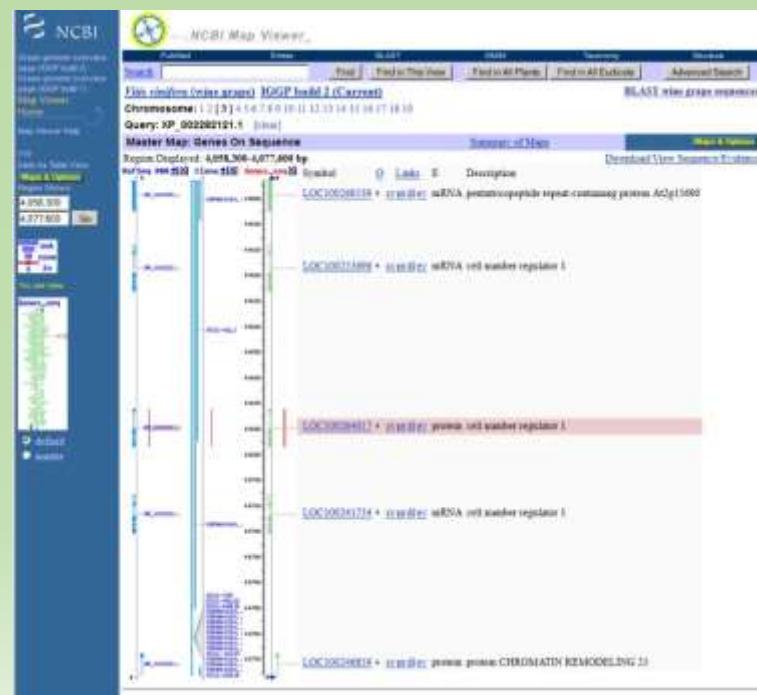
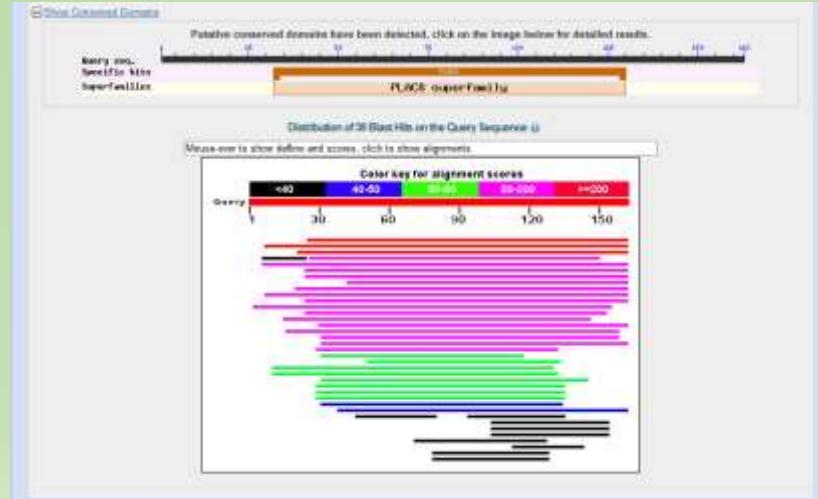
Finding ortholog(s)



Finding ortholog(s)



Finding a BAC



Where has my gene gone? –

<http://www.ncbi.nlm.nih.gov/genome/tools/remap>

Assembly Assembly Clinical Remap Advanced Remap

Genome Information

Source Organism *: Escherichia coli
Source Assembly *: 12X
Target Assembly *: 12X

(or mapping to per a list of available organisms)

Algorithms performed: September 25, 2014, software version: 1.7

First Pass (Reciprocal Best Hits)

| | Total |
|------------------|-------|
| Es_WGS Coverage | 0.936 |
| 12X Coverage | 0.920 |
| Percent Identity | 0.999 |
| Es_WGS Coverage | 0.963 |
| 12X Coverage | 0.045 |
| Percent Identity | 0.998 |

Remapping Options

Minimum ratio of bases that must be remapped: 0.5

Maximum ratio for difference between source length and target length: 0.5

Allow multiple locations to be returned:

Merge Fragments:

Data

Input format: Best Guess
Output format: Same as input

Upload a file: No file selected

OR

Paste data here:

You can paste multiple lines into the text area

Line 1
Es_002240779.2: 48,030 - 53,600

NCBI Resources How To

NCBI Remap Results

Feature remapping information from Es_WGS (GCF_000003745.1) to 12X (GCF_000003745.2)

Summary Data

[Download Summary Data](#)

| ID | Source Features | Remapped Features | Source Intervals | Remapped Intervals |
|----------------|-----------------|-------------------|------------------|--------------------|
| NW_002240779.1 | 1 | 1 | 1 | 1 |

Mapping Report (sample)

[Download Full Mapping Report](#)

| Feature | Src Interval | Remap Interval | Src Location | Src Length | Map Location | Map Length | Coverage |
|---------|--------------|----------------|----------------------|------------|----------------------|------------|----------|
| Line 1 | 1 | 1 | NW_002240779.1:48030 | 5001 | NC_012922.3:15655257 | 5001 | 1.00000 |

Annotation Data

[Download Annotation Data](#)

Genome Workbench Files

[Download Genome Workbench Files](#)

This software/database/presentation is a "United States Government Work" under the terms of the United States Copyright Act. It was written as part of the author's official duties as a United States Government employee and thus cannot be copyrighted. This software/database/presentation is freely available to the public for use. The National Library of Medicine and the U.S. Government have not placed any restriction on its use or reproduction.

NCBI posters

- P0051: New Developments in the NCBI Eukaryotic Genome Annotation Pipeline
- P0791: Genomic Resources for Arabidopsis Research at NCBI
- P1105: From Genes to Genomes
- P1109: dbSNP and dbVAR
- P1112: Improving the Flow of Data to NCBI Sequence Repositories, SRA and GenBank
- P1115: Annotation of Laurasitheria Genomes by NCBI
- P1120: NCBI Resources for Plant Genomics
- P1142: Sequence Viewer and Genome Workbench